融合领域专家信任与相似度的协同过滤推荐算法研究*

谭学清 张 磊 黄翠翠 罗 琳
(武汉大学信息管理学院 武汉 430072)

摘要:【目的】利用领域专家信任和相似度相结合的优势，弥补传统协同过滤推荐算法在推荐准确度以及挖掘长尾商品方面存在的不足。【方法】选取 MovieLens 中稀疏度为 0.9605 的数据集，由评分记录较多的 1 102 个用户对 2 920 部电影的评分记录构成，利用分阶段实验法求得最优专家用户数量及推荐权重系数α值，并结合对比分析法对算法的性能进行评测。【结果】实验结果表明，本算法的推荐结果准确率和覆盖率均受到专家用户数量的影响，且当推荐权重系数为0.6时推荐准确度明显优于传统算法，同时专家用户比例由2%上升至20%时，覆盖率上升了 0.21，说明算法在一定程度上显著提高了推荐系统挖掘长尾商品的能力。【局限】未考虑到不同领域类别之间可能存在的相关性。【结论】该算法能够有效地克服数据稀疏性和冷启动问题，显著提高推荐系统的推荐质量和准确度。

关键词: 个性化推荐 协同过滤 领域专家 相似度
分类号: G354.2

1 引 言

随着个性化推荐系统在 Amazon、豆瓣等网站上的成功应用，加之各种个性化推荐关键技术的发展，个性化推荐的研究已经取得了巨大进展。协同过滤推荐算法是个性化推荐中运用最为成熟的一种推荐算法，然而数据稀疏性和冷启动的问题，在很大程度上影响了协同过滤算法推荐结果的准确度。针对这一问题，现有研究提出了多种解决方法，诸如优化相似度计算方法、引入用户偏好、基于用户聚类等，虽然在解决数据稀疏性和冷启动方面取得了较好的效果，但是由于传统的协同过滤推荐算法依据相似用户给出的推荐结果并不一定能够符合用户的偏好，其推荐结果并没有得到较好的改善。同时，优秀的推荐算法不仅需要准确预测用户的需求，还要帮助用户发现那些感兴趣但并不热门的商品，即对长尾商品的挖掘。

在社交网络中存在这样一类领域专家，他们阅历比较广泛，或者在某一领域有很强的专业性，具有一定的威信，总能给出经验性、指导性的意见。人们也比较倾向于听从领域专家的意见，而这类领域专家能够更客观地对商品进行评分，更少地受到商品的社会化要素影响，对个性化推荐具有积极的推动作用。现实生活中，用户不仅会信任某一类人群，同时也会不信任某一类人群。因此，由领域专家在擅长的领域做出的推荐才更具有可信度，而用户在不同的领域也会选择相信不同的领域专家。

基于领域专家信任的影响，本文认为在推荐过程中应当引入领域专家信任，并提出一种基于领域专家信任与相似度的协同过滤推荐算法。通过对各领域专家用户的挖掘，找到该领域中与目标用户相似的领域专家用户群体，结合领域专家用户与相似用户共同为目标用户进行推荐，既能够优化数据稀疏性等问题，

通讯作者：罗琳，ORCID：0000-0002-4857-9133，E-mail：luo126@126.com。
*本文系国家社会科学基金项目“数字图书馆标签系统的语义挖掘研究”(项目编号：12CTQ003)的研究成果之一。

又有助于对长尾商品进行挖掘。

2 文献回顾

引入信任的协同过滤算法研究中，有学者首先从理论的角度讨论了信任之于协同过滤算法的必要性，例如，Jøsang 等[1]对能够获得信任度和声誉的系统进行综述，认为信任度和声誉应当作为安全机制应用于协同过滤系统中。也有学者通过实验，验证了信任有助于提升协同过滤算法的准确度和覆盖率，例如，Zhang 等[2]和吴应良等[3]将用户间接信任关系和社区特征引入到协同过滤算法中，并验证了这种方式对协同过滤算法的推荐准确度有明显提升。而 Massa 等[4]、Hwang 等[5]、Moradi 等[6]和俞琰等[7]则是将信任度权重引入到协同过滤算法中，替代传统的相似度作为加权的权重，并通过实验验证了基于信任的协同过滤推荐算法可以提高预测准确度和覆盖率。但是这一类研究未能涉及到协同过滤推荐算法面临的数据稀疏性和冷启动的问题，以及推荐算法对长尾商品的挖掘能力。

通过引入信任，在优化协同过滤推荐算法的数据稀疏性及冷启动问题方面，杜永萍等[8]在传统基于用户的协同过滤推荐算法的基础上，引入信任关系计算，设计并构建一个集用户声望信任和用户局部信任的混和信任网络，实现了对协同过滤推荐算法稀疏性问题的优化。Jamali 等[9]构建了基于信任的协同过滤推荐模型 TrustWalker，并依据 Epinions 数据集对模型进行验证，结果显示 TrustWalker 模型在面对稀疏性等问题时比基于信任的推荐方法或协同过滤的推荐方法表现更加优异。Chen 等[10]使用信任与不信任网络识别可靠用户，优化了协同过滤系统中冷启动的问题。Bedi 等[11]基于蚁群整合用户间的动态信任和最优邻居用户的选择，提出了基于信任的蚁群推荐系统，对传统协同过滤算法的稀疏性和冷启动问题进行优化。此外，Lai 等[12]使用评分信任模型和明确的信任准则建立了一种混合的个人信任模型，解决了以往由于评分记录不足而对协同过滤算法推荐结果产生的负面影响。景民昌等[13]提出一种基于专家优先信任的协同过滤推荐算法，并利用 GroupLens 的数据集验证了算法在预测用户评分的精度和成功率上要优于传统的最近邻法。这一类研究验证了信任因素能够解决协同过滤推荐算法面临的数据稀疏性和冷启动的问题，但仍未研究推荐算法对长尾商品挖掘能力的提升。

综上所述，运用信任解决协同过滤推荐算法所面临的数据稀疏性、冷启动、推荐准确度等算法性能问题，已经被多数研究所肯定，但这些研究中聚焦的仅仅是用户之间直接与间接、隐性与显性的信任关系，并未涉及领域专家信任影响，同时未能提升推荐算法对长尾商品的挖掘能力。本文认为引入每一个项目领域中，对该领域贡献评分比较多、比较可靠的权威领域专家，在对项目的评分预测过程中，加入专家信任值权重，能够更好地提高评分预测的准确性，并且有利于挖掘出长尾商品以及缓解冷启动等问题。

3 结合领域专家信任值与相似度的协同过滤推荐算法

实证研究表明加入用户的信任关系可以极大地缓解协同过滤系统不能准确为新用户推荐的问题。用户被鼓励与其他用户相连接来扩充自己的信任网络，但是用户应该选择相信谁是一个很困难的决定。Victor 等[14]指出社交网络中有三类关键人物，分别是：拥有很多知识的专家、处在社交网络中心位置的社交达人、拥有很多评分记录的评分达人，并通过实验验证了这三类关键人物的确能够提高推荐系统的推荐效果。

因而本文提出将社交网络中存在的专家用户的影响力运用于推荐过程中，分领域计算用户的相似度，并结合领域专家和领域相似用户为目标用户进行推荐，即基于领域专家信任与相似度的协同过滤推荐算法(Expert Trust and Similarity CF, ETS-CF)，该算法与基于用户的协同过滤推荐算法对比如图 1 所示。

3.1 识别项目领域专家用户

在现实生活中，专家的意见对人们的决策也有很强的主导作用。如若将社会网络中的人群分为两类：领导者(专家用户)和追随者，追随者往往易于接受领导者的观点，因领导者具有较高的信任度，常常对追随者的喜好具有较大的影响。在协同过滤系统中，评判一个用户是否为专家可以依据该用户的项目评分数量和评分质量，通常情况下，如果用户的评分数量以及质量越高，其成为专家的可能性就越高。那么，在一个领域中需要最先找到评分记录数较多的这一类用户，然后计算每个用户的评分准确度。

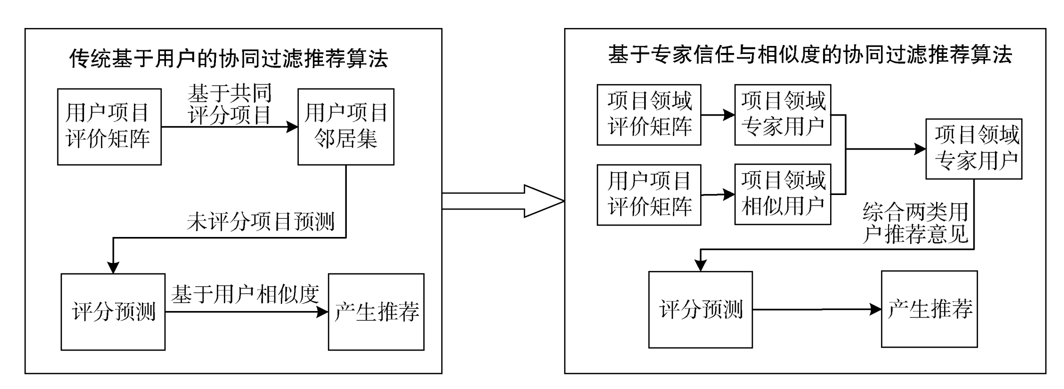


图 1　两类推荐算法对比

在领域 x 中，用户 u 对项目 i 做出的每一次评分，评估其准确度，笔者在 Billsus 等[15]的公式基础上提出以下公式：

$$Pr^x(u)=1-\frac{|r_{u,i}-R^i_{avg}|}{r^i_{max}}\qquad i\in x \qquad (1)$$

其中，$r_{u,i}$ 为用户 u 对项目 i 的评分，R^i_{avg} 为项目 i 的平均评分，r^i_{max} 为项目 i 的最高评分。

关于 R^i_{avg} 的计算，首先需要对偏离评分平均值较远的一些评分数据进行剔除处理，笔者在景民昌等[13]的公式基础上，提出平均评分值计算公式如下：

$$R^i_{avg}=\frac{1}{N}\sum_{u=1}^{N}r_{u,i} \qquad (2)$$

专家的影响度的测量可以从其对项目的累计评分准确度来计算，笔者在 Billsus 等[15]和景民昌等[13]的公式基础上，提出累计评分准确度公式如下：

$$Pr^x(u)=\frac{1}{M_{max}}\times\sum_{u=1}^{M}\left\{1-\frac{|r_{u,i}-R^i_{avg}|}{r^i_{max}}\right\}\qquad i\in x \qquad (3)$$

其中，M_{max} 为在领域 x 中所有用户中评分最多的用户的评分数量，M 为用户 u 在领域 x 中拥有的评分项目总数。

3.2　基于项目领域的用户相似度计算

本文将用户共同评分完成的项目进行领域分类，基于不同的领域计算两个用户的相似度，笔者在 Pearson 相关系数[16]和 Breese 等[17]的公式基础上，提出基于项目领域的相似度计算公式如下：

$$sim^x(u,v)=\frac{\sum_{i\in I}(r_{u,i}-\bar{R}^x_u)(r_{v,i}-\bar{R}^x_v)}{\sqrt{\sum_{i\in I}(r_{u,i}-\bar{R}^x_u)^2}\sqrt{\sum_{i\in I}(r_{v,i}-\bar{R}^x_v)^2}}\qquad I\in x \qquad (4)$$

其中，I 为在项目领域 x 下，用户 u 和用户 v 共同评过分的项目集合。$r_{u,i}$ 为用户 u 对项目 i 的评分值，$\bar{R}^x_u$ 为用户 u 在领域 x 上评分的平均值。$r_{v,i}$，$\bar{R}^x_v$ 同理。

3.3　基于领域专家信任与相似度的推荐算法

在进行评分预测时，首先将专家用户和相似性用户两类用户集合区分开来，然后结合这两类人物的共同推荐意见进行评分预测。在为目标用户选取最近邻居时，不但要选取该领域专家用户，也要选取该领域与目标用户相似的用户。同时，依据专家信任度并利用 Best-k 近邻技术选取邻居集，选取信任度最大的前 k 个用户作为专家用户，并以用户间的相似度值为依据选取相似用户，选取与目标用户相似度值最大的前 n 个用户作为相似用户。结合领域专家信任与相似度的推荐流程如图 2 所示：

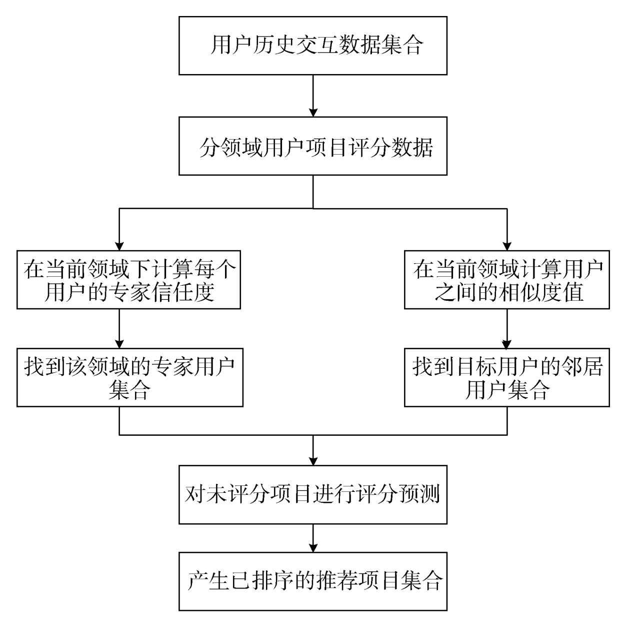


图 2　结合领域专家信任与相似度的推荐流程

针对冷启动用户和非冷启动用户分别有不同的推荐算法，实现如下：

(1) 冷启动用户。由于用户无评分历史或者是在案的评分次数过少，因而无法得知用户感兴趣的领域，则将各领域专家用户的建议推荐给此类用户。笔者在 Ahn[18]的公式基础上，提出评分预测公式如下：

$$P_v^x(i)=\frac{\sum_{u=1}^{k}Pr^x(u)\times r_{u,i}}{\sum_{u=1}^{k}Pr^x(u)} \qquad i\in x \qquad (5)$$

其中，$P_v^x(i)$ 为用户 v 对领域 x 中项目 i 的预测评分值，k 为领域 x 中专家的数量，$Pr^x(u)$ 为专家用户 u 的专家信任度，$r_{u,i}$ 为专家用户对项目 i 的评分值。

(2) 非冷启动用户。依据此领域下专家的意见以及与相似用户的意见，分领域向用户推荐其喜爱的项目。笔者在 Ahn[18]公式基础上，提出评分预测计算公式如下：

$$P_v^x(i)=\bar{R}_v^x+\frac{\alpha\sum_{u=1}^{k}Pr^x(u)\times(r_{u,i}-\bar{R}_u^x)+(1-\alpha)\sum_{u=1}^{k}sim^x(u,v)\times(r_{u,i}-\bar{R}_u^x)}{\alpha\sum_{u=1}^{k}Pr^x(u)+(1-\alpha)\sum_{u=1}^{k}sim^x(u,v)} \qquad i\in x \qquad (6)$$

其中，$\bar{R}_u^x$ 为用户 u 在领域 x 中评分的平均值，$sim^x(u,v)$ 为用户 u 与用户 v 在领域 x 中的相似值。α 是用来调和专家用户与相似用户的评分权重。

4 实验设计

4.1 实验数据集采集与准备

采用 GroupLens 提供的 MovieLens 数据集对算法进行评测。MovieLens 数据集有三种不同容量的数据集，本文采用中等大小数据集并从中截取的部分数据集，并且在实验前，密集化处理了数据集，最终选取的数据集为评分记录较多的 1 102 个用户对 2 920 部电影的评分记录，共包含的 126 784 条电影评分记录，用户的评分包括 5 个等级，稀疏度为 0.9605。此外，通过观察互联网上电影和应用软件的下载规律以及网页点击量与商品销售规律，可以发现虽然热门的物品被大多数用户追捧，但是即使很不热门的物品也会有部分用户喜欢，这就是长尾分布现象。对于 MovieLens 数据集来说，这种长尾现象也存在，如何更好地挖掘长尾商品是推荐系统面临的重要课题。

4.2 实验评价指标

实验采用评分准确度和推荐覆盖率来评测本文提出的推荐算法的性能。对于评分准确度，依旧采取平均绝对误差(Mean Absolute Error, MAE)来衡量。假设在测试集中，I_u 为测试集用户集合，$R_{u,i}$ 为测试集中用户 u 对项目 i 的实际评分，$R_{u,i}^{pre}$ 为在训练集中运用推荐算法得到的用户 u 对项目 i 的预测评分，MAE 公式如下[17, 19]：

$$MAE=\frac{1}{I_u}\sum_{i\in I_u}(|R_{u,i}^{pre}-R_{u,i}|) \qquad (7)$$

除了对推荐算法的准确度进行评估，还进一步评估了推荐算法的覆盖率。覆盖率是算法对长尾商品挖掘效能的体现，覆盖率越高，说明推荐算法越能够将一些经典的不知名的商品推荐给用户。覆盖率的计算公式[20]如下：

$$Coverage=\frac{|U_{u\in U}R(u)|}{|I|} \qquad (8)$$

覆盖率表示最终的推荐列表中包含的物品数占总的项目集合总数的比例。当所有项目基本都被至少推荐给了一个用户，算法的覆盖率比较高，趋于 100%。

4.3 实验步骤

由于条件限制，只对算法进行离线实验，针对算法中提出的领域，实验中将电影的类别信息作为算法中考虑的领域，具体实验步骤如下：

(1) 将经过密集处理完成的 MovieLens 数据集通过随机算法均匀分成 M 份(本次实验中 M 取 10)，选取其中一份作为测试集，其余 M–1 份作为训练集。在训练集上训练得到用户兴趣模型，并对测试集上的用户进行即 10 折交叉验证。为了防止拟合现象对实验结果的干扰，实验将重复进行 5 次，逐次重新随机生成 M 份数据集，将 5 次实验得出的平均值作为最终评测结果。

(2) 寻找每一类电影领域的专家。首先，针对 MovieLens 训练集数据，根据 movies.dat 中每部电影所属的类目将评分数据集 rating.dat 中的评分记录分类，其中电影类目共包括 Action、Drama、Children、Adventure 等 18 个领域。其次，将某一电影领域用户按照评分记录数目从高到低排序，分别计算每个用户

的评分准确度，即专家信任度。最后，将得到的专家信任度值按照从高到低排序，选取该领域最具有专业性的 k 个专家用户，专家信任度作为最终评分预测中的评分权重。

(3) 计算两个用户之间的相似度。首先需要依据电影类别建立项目与用户之间的行为记录表，统计在某一电影类别下，对每部电影有过评分行为的用户，得到记录了同一个电影类别领域所有用户两两之间共同行为项目的总表。随后，针对其他电影领域进行同样的操作，最终得到每一个电影类别下两两用户之间共同的项目行为记录表。

(4) 获得每个电影领域的专家用户并记录该专家用户的信任度，同时记录下所有电影领域每两个用户之间相似度值，分领域为每个用户推荐领域专家以及与其最相似的用户喜欢的 N 个物品(按照评分值排序)。当无法获得在某一个领域与用户相似的用户，则将该领域专家用户推荐的物品给用户。

(5) 对该次推荐进行评测，分别计算推荐系统的准确率以及覆盖率。

5　实验结果与分析

实验主要分为两个部分进行:

第一部分将分为三个阶段对基于领域专家信任与相似度的协同过滤推荐算法进行实验验证。第一个阶段，通过基于用户的协同过滤推荐算法实验确定相似邻居用户数量 k 值。第二个阶段，固定相似邻居用户数量，依据专家用户数量以及推荐权重系数α取值的不同从而导致 MAE 值变化曲线，确定最佳的专家用户数量和推荐权重系数α值。第三个阶段，衡量专家用户数量选取对推荐覆盖率 Coverage 值的影响。

第二部分对三种不同推荐算法的 MAE 值进行对比分析。三种算法分别为: 传统的基于用户的协同过滤推荐算法(USER-CF)、基于专家信任优先的协同过滤推荐算法(EPT-CF)和基于领域专家信任与相似度的协同过滤推荐算法(ETS-CF)。

5.1　实验结果

(1) 传统的协同过滤推荐算法(USER-CF)实验结果以及相似用户数量 k 的确定。

通过 MovieLens 数据集上的离线实验来评测 USER-CF 算法的性能。USER-CF 算法涉及到为每个用户选取多少个最近邻居，也就是 k 值的选取，然后为用户推荐 k 个近邻用户感兴趣的 N 个商品。程序最终为每个用户推荐 N 个物品，由于推荐列表通常都不能很长，因此 N 值固定为 10。通过离线实验评测当相似近邻 k 选取不同的值: 10、20、30、40、50, USER-CF 算法的推荐效果，详细的实验数据如表 1 所示:

表 1　基于用户的协同过滤推荐算法的实验结果

k 值	MAE 值	Coverage 值
10	0.8901	0.6783
20	0.8827	0.5511
30	0.8701	0.4935
40	0.8643	0.4014
50	0.8524	0.3005

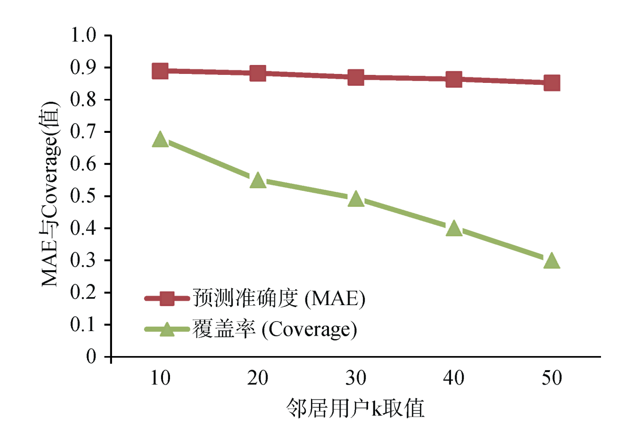


图 3　USER-CF 算法 MAE 与 Coverage 值变化

由图 3 可知，MAE 值随着邻居用户数量的增加而减少，准确率越来越高。Coverage 值随着邻居用户数量增加而减少，用户只会相信与自己最为相似用户的推荐意见，因而推荐的商品数量就趋于减少，也就是推荐系统挖掘长尾的能力随着 k 值的增加而减少。当 k 值为 30 时推荐的性能比较理想，同时兼顾了比较好的准确率与覆盖率。

(2) 专家用户数量选取以及不同权重系数α对推荐准确率的影响。

在实验过程中，针对不同的领域首先选取不同数量的专家用户为普通的电影用户进行推荐。对于每一个领域如何选取专家用户，采取的办法是当某一个领域中用户的专家信任度超过一个预设的阈值时，该用户就成为该领域的专家用户。选取 Action 电影领域的评分记录，根据给出的专家信任度计算公式，计算电

影领域每个用户的专家信任度。用户专家信任度分布如图 4 所示:

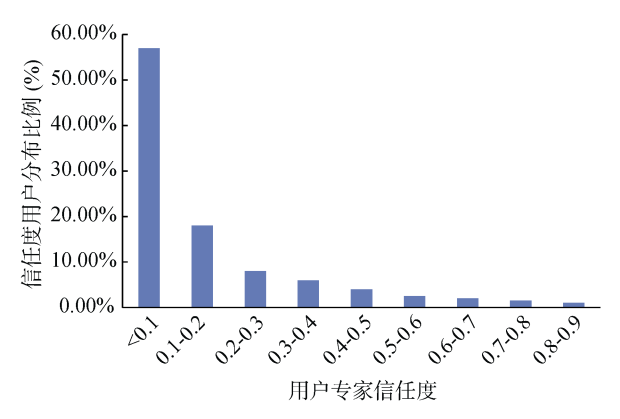


图 4　用户专家信任度分布情况

通过计算每个用户的专家信任度，最大值为 0.803，最小值约为 0.01。由图 4 可知，其中约 58%的用户专家信任度值处于 0.1 以下，约 18%的用户信任度值处于区间[0.1, 0.2)内，剩下 20%左右的用户信任度值高于 0.2。这符合帕列托定律分布，信任度低的用户占了用户总数的绝大部分。

在讨论专家用户的数量对推荐预测准确度的影响时，采取第二种方式，首先将每个用户的相似邻居用户数量固定为 30，每个用户推荐的物品长度固定为 10，按照专家用户比例从 2%-20%变化。同时对权重系数α分别取 0.2、0.4、0.6、0.8 等 4 种不同情形下 MAE 值变化。实验数据如表 2 所示:

表 2　不同专家用户比例与α取值下 MAE 值变化

MAE / 专家用户比例 / α值	2%	5%	8%	11%	14%	17%	20%
0.2	0.8794	0.8421	0.8089	0.7722	0.7524	0.7599	0.7623
0.4	0.8323	0.7990	0.7611	0.7323	0.7044	0.7102	0.7227
0.6	0.8204	0.7807	0.7533	0.7211	0.6977	0.7044	0.7093
0.8	0.8541	0.8327	0.8001	0.7623	0.7274	0.7301	0.7439

如图 5 所示，随着专家数量增多，MAE 值有明显幅度减小，用户喜欢的电影被列入推荐列表中，说明专家数量增多，可以提高推荐模型的准确率。但是当专家用户比例超过 14%, MAE 值又开始有了小幅度回升，这是因为当专家用户占比比较大时，说明存在有的专家用户信任度值比较低，这会造成整体的预测准确度有所下降。

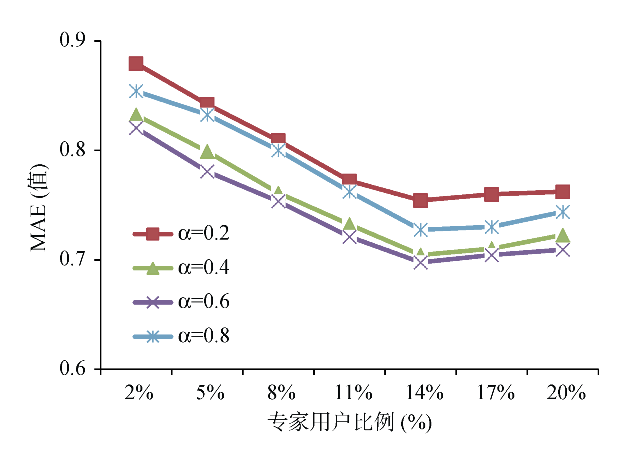


图 5　专家用户比例对推荐 MAE 值的影响

对于模型中专家用户和相似用户权重平衡因子的选取，通过图 5 可以看出：过高或者过低都会导致推荐效果不理想。过高的平衡因子会降低用户自身偏好因素的参与作用，而过多地依赖于专家用户的意见，因忽略用户个人偏好而出现不好的推荐效果。过低的平衡因子会降低领域专家用户的影响力，特别是对于冷启动用户来说，会得不到很好的推荐效果。适当的平衡因子可以提高推荐系统的效果，当平衡因子α取 0.6 时，推荐系统的效果最为理想。因而用户除了听取领域内与其相似用户的意见，也会更加信任领域内权威专家给出的意见。

(3) 专家用户数量的选取对推荐算法覆盖率 Coverage 值的影响。

根据之前的实验结果，在验证专家数量选取对推荐算法覆盖率的影响时，α取 0.6，相似用户数选取为 30 个，为用户推荐的物品长度为 10，专家比例仍旧按照之前的选取规则从 2%-20%变化，实验结果数据如表 3 所示，对应的推荐算法的 Coverage 值变化如图 6 所示。

表 3　不同的专家用户比例对应推荐覆盖率

专家用户比例	Coverage 值
2%	0.6735
5%	0.7025
8%	0.7527
11%	0.7831
14%	0.8204
17%	0.8651
20%	0.8842

由图 6 可知，随着专家数量的增加，推荐的覆盖率不断提高，专家用户可以将某一领域内不热门但是较为经典、价值高的电影推荐给用户观看。这与纯粹的基于用户的协同过滤推荐不同，只会把相似用户认为特别好的电影推荐给目标用户，不利于发掘用户兴趣的多样性和挖掘出长尾商品。因此，本文提出的推荐算法考虑专家用户的影响力，在一定程度上提高了推荐系统挖掘长尾物品的能力。

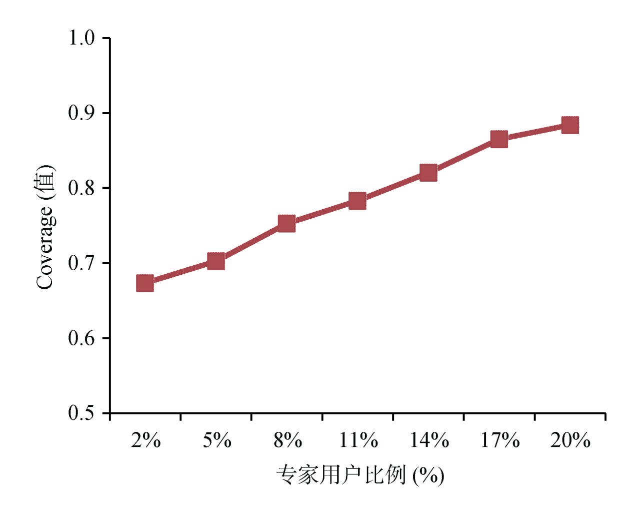


图 6　专家用户数量对推荐覆盖率的影响

5.2　不同推荐算法实验结果比较

选取三种推荐算法，对其推荐准确度方面进行对比。三种算法分别为：传统的基于用户的协同过滤推荐算法(USER-CF)、基于专家信任优先的协同过滤推荐算法(EPT-CF)和基于领域专家信任与相似度的协同过滤推荐算法(ETS-CF)。按照上述实验结果在 ETS-CF 推荐算法实验中取α值为 0.6，设置不同的相似邻居(专家)数量参数 k，三种推荐算法推荐的准确率 MAE 值变化如表 4 所示：

表 4　不同邻居(专家)个数下三种推荐算法 MAE 值变化

MAE值 / 相似(专家)用户个数 / 推荐算法	5	10	15	20	25	30	35	40
USER-CF	0.9043	0.8901	0.8882	0.8827	0.8779	0.8701	0.8660	0.8643
EPT-CF	0.8523	0.8331	0.8201	0.8114	0.8099	0.8071	0.8021	0.7905
ETS-CF	0.8117	0.7993	0.7504	0.7303	0.7111	0.7009	0.7123	0.7201

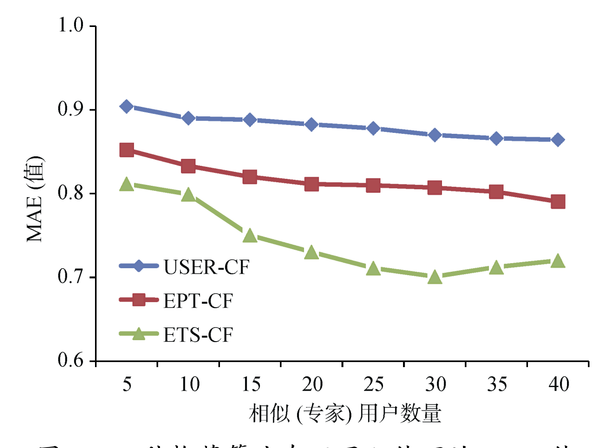


图 7　三种推荐算法在不同 k 值下的 MAE 值

从图 7 对比分析可知，随着相似(专家)用户数量的增多，各算法的 MAE 值都是不断减小，推荐算法的准确度得到提升。但是当 k 值增加到一定程度，MAE 减小的幅度趋于平缓，这是由于与目标用户最为相似的用户推荐意见影响比较大。总体上，基于领域专家信任与相似度的协同过滤推荐算法在预测评分准确度上均优于其余两种算法。特别是，当邻居用户数量比较少时，基于领域专家信任与相似度的协同过滤推荐算法的准确度有明显提升，主要是因为本文提出的算法考虑了专家用户的意见，从专家用户的选取规则中可以看出，专家用户是在某一领域能够给出公正公平见解的一类人群，即便在邻居用户相对较少的情况下，也能为用户做出比较准确的推荐。

6　结　语

针对以往协同过滤算法中存在的问题，本文提出基于领域专家信任与相似度的协同过滤推荐算法，并对该算法利用 MovieLens 数据集进行模拟实验，首先对算法中专家用户数量以及推荐权重系数α进行确定，然后对三类不同的协同过滤推荐算法的实验结果进行对比分析。结果显示，本文提出的基于领域专家信任与相似度的协同过滤推荐算法，推荐结果的准确率和覆盖率均受到专家用户数量的影响，在专家信任度值可接受的范围内，专家用户数量越多，推荐效果越好。同时，通过与基于用户的协同过滤推荐算法、基于专

家信任优先的协同过滤推荐算法进行对比，发现本文提出的算法在推荐准确度方面明显高于前两个算法，并在一定程度上提高了推荐系统挖掘长尾商品的能力。

但是，本文提出的算法仍然存在一定的局限性。该推荐是基于项目领域类别的推荐，并未考虑到不同领域类别之间可能存在的相关性，在实现跨领域精准推荐方面存在一定局限；领域专家用户的评分至关重要，算法启动初期如果不涉及领域专家用户的参与，算法将无法解决“冷启动”的问题，此时和传统算法在“冷启动”方面没有本质差别。

依托于本文算法，笔者未来还将在以下方面做进一步的研究。针对人们日渐增加的情感感知需求，本文只考虑了项目类别情境，而诸如其他的情境因素，例如时间、地点等，也会对推荐结果有显著的影响，那么，如何有效地将其他情境因素运用于个性化推荐中在未来是一个重要的方向。在现实生活中，网站的数据量将会远远大于 MovieLens 数据集，数据关系也更加复杂，如何优化该推荐算法的可扩展性及复杂度等问题，是一个重要的研究课题。

参考文献：

[1] Jøsang A, Ismail R, Boyd C. A Survey of Trust and Reputation Systems for Online Service Provision [J]. Decision Support Systems, 2006, 43(2): 618-644.

[2] Zhang B, Huang Z, Yu J, et al. Trust Computation for Multiple Routes Recommendation in Social Network Sites [J]. Security & Communication Networks, 2005, 12(12): 159-174.

[3] 吴应良，姚怀栋，李成安. 一种引入间接信任关系的改进协同过滤推荐算法[J]. 现代图书情报技术，2015(9): 38-45. (Wu Yingliang, Yao Huaidong, Li Cheng’an. An Improved Collaborative Filtering Recommendation Algorithm with Indirect Trust Relationship [J]. New Technology of Library and Information Service, 2015(9): 38-45.)

[4] Massa P, Avesani P. Trust-Aware Recommender Systems[C]. In: Proceedings of the 1st ACM Conference on Recommender Systems. 2007.

[5] Hwang C S, Chen Y P. Using Trust in Collaborative Filtering Recommendation [A]. // New Trends in Applied Artificial Intelligence[M]. Springer Berlin Heidelberg, 2007: 1052-1060.

[6] Moradi P, Ahmadian S. A Reliability-Based Recommendation Method to Improve Trust-Aware Recommender Systems[J]. Expert Systems with Applications, 2015, 42(21): 7386-7398.

[7] 俞琰，邱广华. 融合社会网络的协同过滤推荐算法研究[J]. 现代图书情报技术, 2012(6): 54-59. (Yu Yan, Qiu Guanghua. Research on Collaborative Filtering Recommendation Algorithm by Fusing Social Network [J]. New Technology of Library and Information Service, 2012(6): 54-59.)

[8] 杜永萍，黄亮，何明. 融合信任计算的协同过滤推荐方法[J]. 模式识别与人工智能，2014, 27(5): 417-425. (Du Yongping, Huang Liang, He Ming. Collaborative Filteration Recommendation Algorithm Based on Trust Computation [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(5): 417-425.)

[9] Jamali M, Ester M. TrustWalker: A Random Walk Model for Combining Trust-Based and Item-Based Recommendation [C]. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009: 397-406.

[10] Chen C C, Wan Y H, Chung M C, et al. An Effective Recommendation Method for Cold Start New Users Using Trust and Distrust Networks [J]. Information Sciences, 2013, 224(2): 19-36.

[11] Bedi P, Sharma R. Trust Based Recommender System Using Ant Colony for Trust Computation [J]. Expert Systems with Applications, 2012, 39(1): 1183-1190.

[12] Lai C H, Liu D R, Lin C S. Novel Personal and Group-Based Trust Models in Collaborative Filtering for Document Recommendation [J]. Information Sciences, 2013, 239(4): 31-49.

[13] 景民昌，唐弟官，于迎辉. 基于专家信任优先的协同过滤推荐算法[J]. 图书情报工作，2012, 56(11): 105-108. (Jing Minchang, Tang Diguan, Yu Yinghui. A Recommending Method Based on Expert Prior Trust in Collaborative Filtering [J]. Library and Information Service, 2012, 56(11): 105-108.)

[14] Victor P, Cornelis C, De Cock M, et al. Key Figure Impact in Trust-enhanced Recommender Systems [J]. AI Communications, 2008, 21(2-3): 127-143.

[15] Billsus D, Pazzani M J. Learning Collaborative Information Filters [C] In: Proceedings of the 15th International Conference on Machine Learning. 1998.

[16] Rodgers J L, Nicewander W A. Thirteen Ways to Look at the Correlation Coefficient [J]. American Statistician, 1988, 42(1): 59-66.

[17] Breese J S, Heckerman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering [C]. In:

Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. 2013: 43-52.
[18] Ahn H J. A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-Starting Problem [J]. Information Sciences, 2008, 178(1): 37-51.
[19] Shardanand U, Maes P. Social Information Filtering: Algorithms for Automating "Word of Mouth" [C]. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1995.
[20] Herlocker J L. Evaluating Collaborative Filtering Recommender Systems[J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53.

作者贡献声明:

谭学清: 提出研究思路, 设计研究方案, 修改论文;
张磊, 黄翠翠: 进行实验, 采集、清洗和分析数据, 起草论文;
罗琳: 研究方案的修改, 论文最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。
[1] 张磊, 黄翠翠. Movielens.xlsx. 基于领域专家信任与相似度的协同过滤推荐算法研究文章原始样本数据.
[2] 张磊, 黄翠翠. observedata.xlsx. 基于领域专家信任与相似度的协同过滤推荐算法研究文章观察数据.
[3] 张磊, 黄翠翠. cfusingdata.xlsx. 基于领域专家信任与相似度的协同过滤推荐算法研究文章样本数据.

收稿日期: 2016-04-04
收修改稿日期: 2016-05-23

A Collaborative Filtering and Recommendation Algorithm Using Trust of Domain-Experts and Similarity

Tan Xueqing　Zhang Lei　Huang Cuicui　Luo Lin
(School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract: [**Objective**] This paper tries to improve the performance of traditional collaborative filtering and recommendation algorithm. [**Methods**] We used the MovieLens dataset to evaluate the proposed algorithm. First, chose datasets with sparse degree of 0.9605, which included scoring records of 1,102 users for 2,920 movies. Second, identified the optimal number of expert users and recommended weight coefficient alpha value with series of experiments. Finally, evaluated the algorithm's performance with comparative method. [**Results**] The precision of the algorithm were influenced by the expert users. When the recommended weight coefficient value was 0.6, the precision of the new algorithm was better than the traditional ones. Once the propotion of expert users increased from 2% to 20%, the coverage value increased by 0.21. Thus, the new algorithm could analyze the long tail goods more effectively. [**Limitations**] We did not take into account the possible correlation among different categories. [**Conclusions**] The proposed algorithm could effectively solve the data sparsity and cold start issues, which significantly improve the performance of the recommendation system.

Keywords: Personalized recommendation　Collaborative filtering　Domain-Expert　Similarity